Mining Interest In Online Shoppers' Data: An Association Rule Mining Approach

Michael Adu Kwarteng, Michal Pilik, Eva Jurickova

Tomas Bata University in Zlin, Faculty of Management and Economics,
nam. T. G. Masaryka 5555, 760 01 Zlin, Czech Republic
kwarteng@fame.utb.cz, pilik@fame.utb.cz, jurickova@fame.utb.cz

Abstract: Online shopping, as a form of e-commerce, is not nearing extinction anytime soon. As the interplay between shoppers and vendors continues to grow in the midst of complex transactional data, extracting knowledge from the data has become imperative. In view of this, this paper explores the use of the association rule mining technique to glean relevant information from such shopper-vendor interactions. In particular, this paper looks at some of the unusual, frequent relationships existing between online shoppers on one hand, and vendors on the other hand in the Czech Republic. The results revealed with higher confidence values the following: (1) there is a strong association between criteria for buying items on the Internet and information gathered before initiating an online transaction; (2) a sizable number of online customers engage in online shopping because of the price attached to the product in question; and (3) a greater proportion of online customers engage in online transactions through specialized e-shops. The work provides general insights into how shopper-vendor transactional data can be explored.

Keywords: E-commerce; online shopping; consumer buying behavior; association rule mining

1   Introduction

Online shopping contributes to the massive volumes of data in e-commerce. The emergence of the Internet and other technologies have ensured that these large transactional data are stored, processed, and analyzed with speed. In fact, the exponential increase in online shopping is clear evidence of the growth in electronic buying and selling. The usages of several digital modes by many companies worldwide have contributed by and large to the resilient electronic platform. For example, marketers are now able to reach many prospective and existing consumers at a minimal cost, hence aiding them to achieve their set targets [1, 2]. Similarly, most companies that operate via the electronic medium are able to take hold of a large share of clients at a minimal cost. This stems from the introduction of marketing tools such as viral or email marketing in addition to

special digital offers and periodic promotions firmly supported by other multimedia resources [3]. According to Min and Wolfinbarger [4] and Brown [5] resulting in huge sales and profits annually. Ahluwalia, Hughes, and Midha [6] and Lennon, Johnson, and Lee [7] cited a classic example indicative of huge profits that emanate from large annual sales with specifics on the infamous "Black Friday" sales. Indeed, they observed the fast pace that companies take advantage of the Internet business and online shopping, with less cost of an input to rake in so much profit during annual festive occasions around the globe.

Various studies have sought to confirm the conception that online shopping is gaining roots as a new positive feature for sales increase and retention of customers [8, 9, 10, 11]. In line with this trend, many authors [12, 13, 14] also offered various opinions on some of the ways online shopping can become important to most companies. They contended that this requires analyzing the value proposition for customers of these companies in order to acquire the needed data from digital-savvy consumers for research purposes. In these circumstances, well-known multinational companies like Coca Cola, Amazon, eBay, and Proctor and Gamble (P&G), among others, have been regarded as perfect examples of companies that have—by virtue of Internet technologies—fruitfully shifted from brick and mortar strategies to brick and click techniques [15, 16].

Eurostat [17] further reported on the scale or extent that consumers have redirected their concentration from the brick and mortar style of shopping to the brick and click style. The study indicated that consumers (a) appreciate the convenience of being able to shop anytime and anywhere; (b) applauds the ease in acquiring information from a broader range of products; and (c) are satisfied with the ability to compare prices and share their opinion of goods with other consumers. In respect of this, [18] posited that online shopper-vendor retail sales have shot up to an estimated USD$840 billion in 2014. This exceeded the USD$695 billion in sales recorded in the preceding year of 2013, and sales were projected to rise to USD$1.506 trillion in the year 2018. Earlier, EMarketer [19] revealed that global online sales were expected to garner a whopping sum of USD$1.915 trillion; this would be accompanied with a sales growth of 23.7% in 2016. The study further revealed a projected increase of USD$4.058 trillion with such a corresponding increase, making up 14.6% of its total retail spending in the year 2020.

In spite of all these benefits attained by e-vendors as a result of the introduction of online shopping, online providers need to keep track of their regular e-shoppers. In a broader context, this brings to the fore how online companies and their respective customers interact by extending new resolutions in terms of customer relationships and customer loyalty [20]. To analyze both geographic and demographic patterns of online buying behavior, this paper uses a data mining algorithm—association rule mining—to mine relevant patterns of online shopper-vendor interactions from a large data set in the Czech Republic. The paper fills the gap in the literature about online shopper-vendor interactions, and primarily aims

at widening the general concept of online buying behavior. The findings of this study are of great significance to e-vendors in achieving value proposition for their customers. In addition to e-commerce or e-vendors, industry relevance, the study will thus add to the scant research output on association rules purposed for online buying behavior, and subsequently shed light on online customers' preferences in e-stores associated with the Czech Republic. As a result, the goals of this study are to (a) analyze the reasons and rationales that influence customers to buy via the Internet; (b) determine the criteria embarked upon by customers in initiating online transactions; (c) access the requisite information needed to make online transactions; and (d) identify the particular type of online shops consumers prefer to use, coupled with the specific products most commonly purchased online.

The rest of the paper is structured as follows. Section 2 provides a review of the literature and the theoretical background on this topic. Section 3 briefly explains the concept of association rule analysis and its related applications, accompanied by research questions used in this paper. Section 4 explains research questions and the methodology, including the samples and procedures used. A case study of the association rule analysis approach is conducted using a large data set acquired from respondents in the Czech Republic. The main results are presented and discussed in Section 5, followed by a summary and conclusions.

2 Synopsis of Online Buying Behavior in the Czech Republic

Online shopping in the Czech Republic dates back to the second half of the 1990s, and was initiated by a group of IT specialist who turned American-styled ideologies into a reality [21]. According to Redakce [21], the turn of events in the Czech Republic regarding sales, retailers, and marketing in general became "saturated" in the millennium year, with customers churning away from their usual brick and mortar stores because of the conditions meted out to them. As a matter of fact, the startling revelation of Internet retail took turns from this era, since wholesale and retail had been entrenched. It was easier for retailers to use their existing stores coupled with their vast array of experience in computers to penetrate Internet retail as earlier envisaged. The initial establishment of Internet retail saw Alza, Czech computers, and Mironet entering the game of Internet retail in the Czech Republic [21].

In mid-2012, the Czech Statistical Service reported that almost 2.6 million persons, representing 31% of the entire population in the Czech Republic and 44% of the total number of Internet users in the country, claimed they had shopped online in 2012 [22]. According to Czech News Agency [23], the number of online

shops in the Czech Republic increased from 16% to 37% over the last two years. As part of the rapid penetration of Czech customers in online transactions, a study on e-commerce is carried out regularly to monitor the inflow and outflow of customers in collaboration with notable companies from the Czech Republic, such as Gemius, Centrum Holdings, and Seznam.cz. The rationale of the annual study is to probe into the distinction between Czech Internet users and the tendency of these users to shop online. Some of the valuable facts retrieved from this study are not limited to the fact that Czech Internet users are driven to shopping each and every time they are online. Women were also seen as the most inclined online shoppers in terms of gender distribution; approximately 60% or more take advantage of Internet shopping in the Czech Republic [22]. Again, the report indicated that Czech Internet users who have already transacted business online continue to grow unprecedentedly. Moreover, a sizable proportion of Internet users start shopping for something new online. Yet, the report further opines that almost three-quarters of Czech Internet users have not finished a transaction or purchase through the Internet at least once. It is within this context that we use association rule analysis to explore online shopper-vendor interactions in the Czech Republic.

The setting of this exploratory study, the Czech Republic, appears to stand tall among other Central European sister countries in terms of infrastructure and market size as well as in the total global position of the online retail market (stores). This gives credence to the study conducted by Hutchings, Best and Mahmuti [24] that ranked the Czech Republic as 33rd in the world for online buying and 20th among the list of Internet retail sales per capita of USD$217 million in 2012. Again, a report by eCommerce news [25] indicated that in the Czech Republic there has been an increase in online transactions to more than 20% with an amount of CZK$25 billion, which is equivalent to €$925 million. The report further stated that Czech customers spent €$3 billion in their local online stores in 2016; this was estimated to double by 15% in the subsequent year.

3   Association Rules

The association rule mining (AR) approach is one aspect of a data mining algorithm that is used for market basket analysis; it is specifically useful and viable in retrieving and discovering interesting relationships concealed in a large data set [26]. The impact of AR methodology, helps retailers analyze a large data set, among other things, with the intention of learning more about the purchasing pattern of customers [27, 28]. Over the years, this algorithm has been widely used by astute researchers in academia and the industry in general, specifically in the area of marketing. However, the full-fledged application of AR methodology could be attributed to Agrawal, Imielinski and Swami [29]; later works by

Agrawal and Srikant [30] assisted in shaping the algorithm into a well known concept for utilization in academia. Other renowned scholars [31, 32] cannot be left out as championing the cause of AR methodology into the mainstream research seen today.

The method has been curbed to fit into diverse data forms feasible in different and multiple scenarios. As part of its standard usage and application, association rule analysis has been used in many academic disciplines in a variety of areas within academia and industry as earlier indicated. For example, Becquet, Blachon, Jeudy, Boulicaut, and Gandrillon [33] used association rule mining in the application of gene-expression data analysis to detect sets of genes whose expressions were interrelated, while Borgelt and Berthold [34] also made use of association rules to analyze fragments embedded in molecules that assist to discriminate between diverse classes of molecules. AR has in many ways resulted in other modifications apart from the Apriori algorithm. These include the Frequent-Pattern Growth algorithm and the Eclat algorithm [35, 36, 37]. However, these algorithms are categorized into two stages before their application. The reliability, accuracy, and strength of AR mining produced in many research areas and academic disciplines gives credence to its relevance and accurate usage in any research endeavor [38].

Conventionally, AR mining is denoted by $A \geq B$, showing when the two products (A and B) were purchased. A then becomes the antecedent, while B is denoted as the consequent, indicating that A influenced the purchase of B.

4   Research Questions and Methodology

Though in extant studies the authors have positioned online shopper-vendor interactions [39, 40, 41, 42], the pattern of online shopping behavior has been silent. In other words, using a data mining algorithm (association rule mining) to mine the relevant patterns of online shopper data as earlier stated is a concern. To this end, these research questions are formulated to guide the study:

1. What are the reasons that cause the customer to start buying on the Internet?
2. What are the criteria for buying items on the Internet?
3. What information is needed or gathered before initiating online transactions?
4. Which online shops do customers prefer, based on their previous online purchases?

This study was conducted on a notion of a probability sampling method - pseudo simple random sampling technique. Data collection was centered on self-administered questionnaires. In particular, we collected data from different

geographic regions in the Czech Republic, and focused on people who have shopped online before or have yet to shop online. The demographics and geographic location of respondents were also pertinent to the research; this necessitated the inclusion of respondents from the biggest cities across the Czech Republic—notably Prague, Brno, and Ostrava. The research was carried out during 2015 and 2016. To execute the agenda of our study, both RapidMiner data mining software (an open source data mining software) and Adindoft Xlstart (2014) software were used for the whole analysis. We disbursed 1,803 questionnaires to respondents for the study, of which 1,602 (88%) were accurately filled out and returned to us for evaluation. The rules generated from our data were coined from the questions presented to online shoppers (the respondents) from the Czech Republic, and generally focused on determining the reasons that propel e-shoppers to start buying on the Internet, conditions for buying items on the Internet, and the information gathered from customers before initiation of online transactions. It can be seen from Table 1 that most of our antecedents were mainly generated and extracted toward "specialized online shops" and "price" as the main consequents for our discussion. Table 1 provides a vivid analysis of the associations retrieved from e-shoppers in the Czech Republic and shows the frequencies and percentages for some demographic factors.

Table 1

Sample demographics

Demographic considerations of the respondents		Frequencies (N)	Percentages (%)
Gender	Male	665	41.5
	Female	937	58.5
Age	16–24	418	26.1
	25–34	392	24.5
	35–44	270	16.9
	45–54	203	12.7
	55–64	170	10.6
	64+	149	9.3
Study level	Primary education	112	7.0
	A Secondary school without graduating	151	9.4
	Secondary school with graduation	572	35.7
	University degree	767	48.9
Level of Internet literacy	Beginner	243	15.1
	Common user	806	50.3
	Advanced user	391	24.4
	Professional	162	10.1

Next, we applied the association rule analysis technique to analyze and learn the patterns inherent in the large data set. This helped discover the associations connected with respondents' choices of embarking on online transactions. Association rule analysis follows the systematic procedures or steps below. The adopted methodology is generally focused on the terminologies listed below and based on the Apriori algorithm which was used in this study. The algorithm works by expounding on the following mathematical theory of Agrawal, Imielinski and Swami [23]:

Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of n binary attributes called items.   (1)

Let $D = \{t_1, t_2, \ldots t_m\}$ be a set of transactions referred to as the database.   (2)

Each *transaction* in D has a unique transaction ID and contains a subset of the items in I.

AR mining assumes the form A ≥ B, where A and B are item sets and are satisfying $(A, B \subseteq I)$; (e.g., {Milk, Diaper} → {Beer}). The right-hand side of the rule (A) is termed the antecedent or the premise and the left-hand side the consequent or the conclusion.

In extracting or generating rules for analysis, we selected the variables of interest to mine the relevant relationship patterns. Variables of interest were converted into a transactional data format to aid association rule mining. Three important metrics underpin the foundation and deployment of association rules. These are measures of "support", "confidence", and "lift":

1. Support: The support of an association rule A ≥ B is the percentage (%) of transactions in the database that contain both A and B. Mathematically, Support $= frq\,A, B/N$, where N is the number of transactions.

2. Confidence: The confidence (Φ) for an association rule A ≥ B is the proportion attributed to the number of transactions that is made up of A U B assigned to the number of transactions that comprises A. This is shown mathematically as: Confidence (Φ) $= \left(frq[A, B]\right)/\left(frq[A]\right)$.

3. Lift: The lift of an association rule is the fraction of the support of (A, B) to the support of (A) and the support of (B). In practice, the lift looks at the left-hand side rule and finds the percentage of chance of (B) appearing would increase: Lift $=$ $\left(support[A, B]\right)/\left(\left[support\{A\}\right] * \left[support\{B\}\right]\right)$

5 Empirical Findings and Results

As earlier indicated, association rules are presented in antecedent A ≥ B format. The two main metrics used for rule evaluation are confidence and support. Support is the fraction of transactions that contain both the antecedent and the consequent, whereas confidence measures how often items in the consequent appear in transactions that contain the antecedent. It must be noted that when associations between three or more attributes are found (e.g., {Milk, Diaper} → {Beer}), the confidence percentages are computed based on the two attributes being found in the third. Other metrics such as lift, Laplace, gain, and so on, are additional indicators that demonstrate the strength of the rules' relationships [43]. Table 2 and Table 3 present some preliminary findings of association rules for our large data set from the Czech Republic.

Table 2
Binary attributes of the data set toward specialized shops

Rules	Antecedent (A)	Consequent (B)
#1	Price, clothes, shoes	Specialized online shops
#2	Clarity and menu navigation	Specialized online shops
#3	Convenience (home delivery), clothes, shoes	Specialized online shops
#4	Information from friends, colleagues (word of mouth), clarity, and menu navigation	Specialized online shops
#5	Price, Internet discussions	Specialized online shops
#6	Price, convenience (home delivery), electronics	Specialized online shops
#7	Electronics, clarity, and menu navigation	Specialized online shops
#8	Information from friends, colleagues (word of mouth), electronics	Specialized online shops

Table 3
Binary attributes of the data set toward price

Rules	Antecedent (A)	Consequent (B)
#1	Electronics, clarity, and menu navigation	Price
#2	Specialized online shops (e.g., Alza), official reviews	Price
#3	Specialized online shops (e.g., Alza), information from friends, colleagues (word of mouth), electronics	Price
#4	Information from friends, colleagues (word of mouth), clarity, and menu navigation	Price

#5	Convenience (home delivery), terms of delivery	Price
#6	Electronics, terms of delivery	Price

Tables 2 and 3 provide both antecedents (premises) and the consequents (conclusions) of some of the online shopper-vendor interactions in the Czech Republic as our case study for the evaluation.

This section elaborates some of the main findings retrieved from respondents. The association rules extracted from online shopper-vendor interactions in the Czech Republic were centered on specialized online shops (e.g., Alza, Heureka, etc.) and the price of purchasing a particular product via the web or online. Specialized online shops recorded the highest rules extracted from our data followed by price as the consequent. The implication is that Czech online customers are keen on specialized online shops anytime they pause to embark on online transactions, while the price in question becomes the second option of initiating online transactions. Detailed analyses are discussed in the next chapter, given the confidence and support of the extracted rules.

5.1 Association Rule Analysis Generated and Extracted (Price)

Table 4

Association rules extracted from online shopper-vendor interactions

Rules	Association rule	Confidence
#1	{Electronics, clarity, and menu navigation} ⟶ {Price}	81%
#2	{Specialized online shops (e.g., Alza), official reviews} ⟶ {Price}	81%
#3	{Specialized online shops (e.g., Alza), information from friends, colleagues (word of mouth), electronics} ⟶ {Price}	81%
#4	{Specialized online shops (e.g., Alza), quick goods comparison} ⟶ {Price}	81%
#5	{Specialized online shops (e.g., Alza), Internet discussions} ⟶ {Price}	83%
#6	{Convenience (home delivery), terms of delivery} ⟶ {Price}	85%

#7	{Specialized online shops (e.g., Alza), terms of delivery} ⟶ {Price}	85%
#8	{Information from friends, colleagues (word of mouth), terms of delivery} ⟶ {Price}	85%
#9	{Electronics, terms of delivery} ⟶ {Price}	86%

Table 4 presents rules generated with a confidence of more than 80%. The findings shed more light on price as the consequent of our analysis. Rule #8 indicated that when online Czech customers pause during their transactions to ask for more information from friends, colleagues or about the terms of delivery, the overriding factor of consideration they looked out for was price, which is accompanied by a confidence of 85%, indicating a higher association between the antecedent and the consequent. Likewise, Rule #9 implied that when customers are buying electronics and thinking of the terms of delivery, the most important factor of influence is price, representing a higher confidence of 86% of influence of price as a deciding factor. In the same vein, when online customers decide to make a decision to purchase electronic gadgets or equipment, and tend to seek clarity and probe into the menu on the website, the prevailing factor is attributed to the price of the said electronic appliance or gadget with 81% confidence.

Additionally, when customers, channel their attention to specialized online shops (e.g., Alza, Heureka, etc.) in the Czech Republic with a higher emphasis on official review of the said product (as Rule #2 stipulated), then it is as a result of the product in question. This scenario, however, recorded 81% confidence similarly to Rule #4 which indicated that online Czech customers hold the view of rationalization through the web by comparing the various prices for a similar product before purchases are made. This means that Czech customers are more particular about the price of the products on the web when the product is basically situated in a specialized online shop. Hence, they tend to seek more information from previous users and subsequently check other online shops to confirm the price range.

Alternatively, in both Rule #3 and Rule #5 involving the bid for online shoppers to consider purchasing through specialized online shopping, while shoppers seek information from Internet discussions, friends, or word of mouth when attempting to purchase electronic appliances, the leading factor was price, with 81% and 83% confidence, respectively, for both rules. This means that online Czech customers tend to consider the price of an electronic appliance through a specialized shop in an attempt to initiate a purchase. Again, Rule #6 presented quite different scenarios from all the associations generated from our data set. Rule #5 indicated 85% confidence from Czech online customers who intend to make purchases on the web with the notion of convenience (home delivery) as well as terms of

delivery being the guiding principle of the price that needs to be paid for the product in question. This rule was supported by 85% confidence indicating that customers are concerned about how the goods will get to them, and the way and manner these goods will be delivered is embedded in the price of the goods.

5.2   Association Rule Analysis Generated and Extracted (Specialized Online)

This section presents the results calculated for specialized online shops as a consequence and the antecedents or premise. The extracted variables generated in tandem with the consequent (specialized online shops) were price, information from friends, and convenience (home delivery, electronic appliances) as a background. These are discussed within the subsequent chapter with figures provided for clarification.
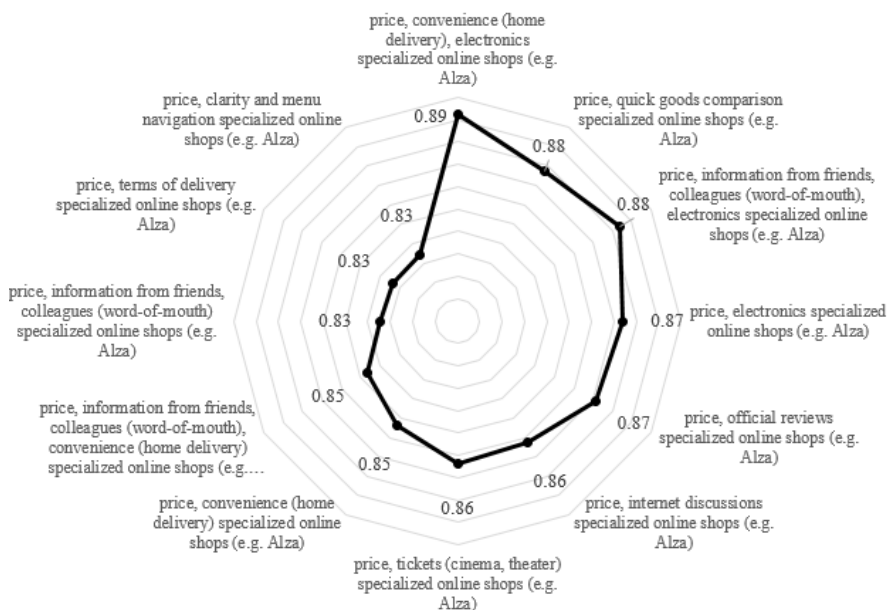

Figure 1
Association rules of price and specialized online shops

The association rule extracted from Figure 1 provides confidence and support of rules embedded in price and specialized online shops in the Czech Republic. As shown in Figure 1; price, convenience (home delivery), electronics, specialized online shops with the highest confidence of 89%. This means that when online customers are focused on purchasing an electronic appliance via a specialized online shop in the Czech Republic, they look out for the price and how the said

electronic appliance will get to them. Second to the highest rule (recorded with 88% confidence), online customers in the Czech Republic tend to seek clarification from friends and to subsequently do a quick price comparison anytime they decide to purchase a product from specialized shops across the Czech Republic. Price was extracted with a higher confidence as a deciding factor of online customers to embark upon specialized online shops; see Figure 1. However, the lowest confidence extracted from the figure stood at 83%, which was centered on information from friends, clarity and navigation, terms of delivery associated with specialized online shops, and price as a deciding factor for online customers. This means that online customers seek clarity and first-hand information from colleagues and also on how the products yet to be purchased (price) will get to them from the specialized online shops.
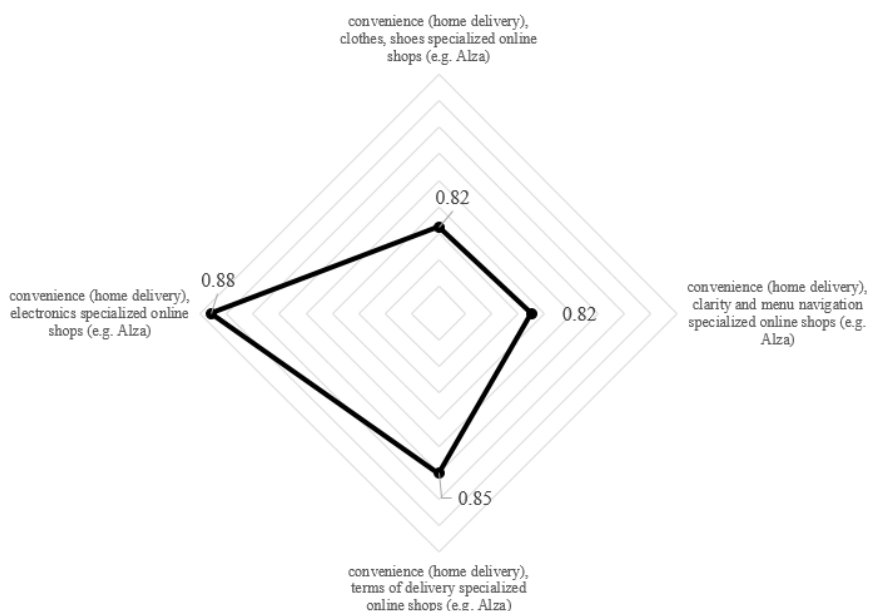

Figure 2
Association rules convenience and specialized online shops

Figure 2 presents an extracted rule from convenience (home delivery) with respect to specialized online shops in the Czech Republic. The highest rule from Figure 2 recorded 88% confidence, convenience (home delivery), electronic specialized online shops. In the event that a customer is targeting an electronic appliance from a specialized online shop, the tendency for the particular electronic appliance to be purchased stems from the fact that the product will be delivered accurately without any hindrance. The second rule generated stood at a confidence level of 85% showing that convenience, terms of delivery are a domain of specialized online shops. This is attributed to the fact that online customers, channel their

attention toward specialized shops when the product they wish to purchase will be delivered without obstruction, and the terms of delivery are well negotiated. Alternatively, the last two rules regarding specialized shop engagement in the Czech Republic present an interesting rule that indicates online customers tend to adopt convenience, and clarity and navigation when they intend to purchase goods like clothes, shoes, and so on, with 82% confidence in those two rules extracted. It is an obvious fact that online customers will seek more clarity when they intend to purchase goods of such caliber because of the size, color, and so on. Since these qualities cannot be experienced online like brick and mortar stores, it becomes vital for the online customer to get to the specialty online shops that deal with a specific kind of item.
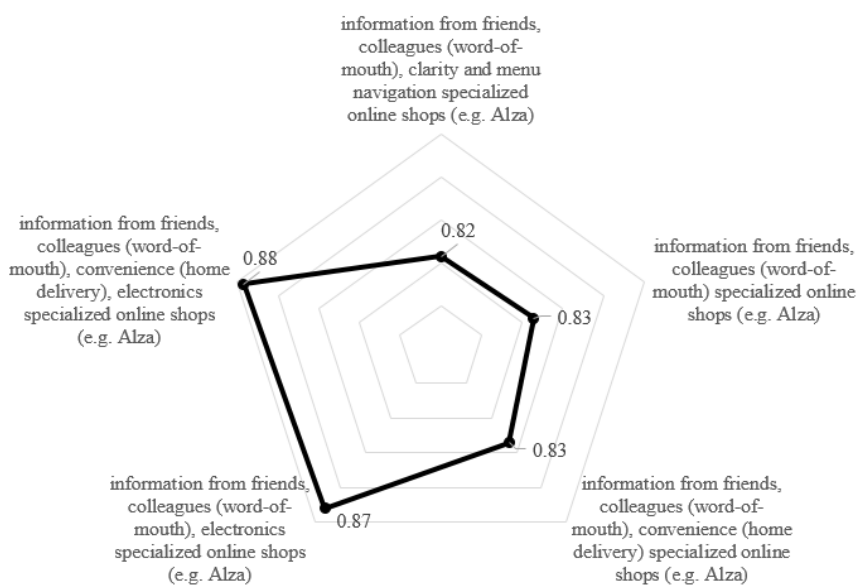

Figure 3
Association information from friends and specialized online shops

The association rules extracted from information from friends against the inclination to shop at specialized online shops in the Czech Republic are presented in this section. The highest confidence retrieved from our data stood at 88%. The extracted association centered on information from friends, colleagues (word of mouth), and clarity and navigation, with the specialized online shop as the consequent. This implies that in the quest of online customers engaging in specialized online shops, the dominant factor is seeking information both from friends and at the site before initiating an online purchase. Again, in the event of online customers planning to purchase any electronic appliances via specialized online shops, the deciding factor is feedback from friends who have already purchased the said product, with 87% confidence. Interestingly, there were two

extracted rules that recorded the same confidence of 83%, with convenience as the only distinguishing factor. As previously indicated, this implies that online customers inquire more with their friends or colleagues before embarking upon specialized online shops, and also look out for the convenience of the product they intend to purchase. In contrast to the previous claims made by respondents as to the association between information from friends and the need to adopt specialized online shops, it can be seen from Figure 3 that the clarity of the site as well as colleagues (word of mouth) plays a pivotal role for the online customer in initiating specialized online shopping; this stood at 82%, indicating the lowest confidence from the respondents.
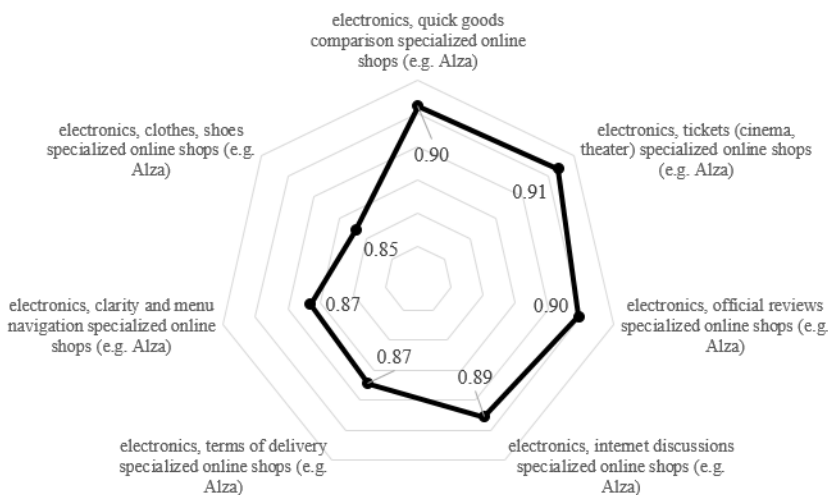

Figure 4
Association rules electronics and specialized online shops

Figure 4 represents the associations extracted from our data. It can be noted that the tendency for online customers to pause to buy electronic gadgets as well as tickets to cinemas or theatres in a specialized online shop cannot be overemphasized. This assertion recorded a confidence of 91% as the highest of all confidence retrieved from the association between electronic appliances and specialized online shops. This means that online Czech customers tend to locate specialized online shops anytime they think of buying tickets for cinema or theatre and also any electronic gadget. The second rule extracted from the above scenarios centered on convenience, official reviews, and Internet discussions by online customers when they embark on specialized online shops to purchase any electronic appliances. These rules recorded both 90% and 89% confidence; that is to say, online Czech customers are more particular about how their electronic appliances will get to them and also want to delve into what others are saying about the said electronic appliances or gadgets.

Conclusions

This study is aimed at applying a data mining technique, to evaluate and analyze online shopper-vendor interactions, with the Czech Republic as our case study. Online transactions are evolving, and are both a blessing and a challenge for both the e-vendor and the e-shopper. To ensure smooth operations, research must generate and extract relevant information and transactional data from online customers in the Czech Republic. The association rule mining approach was used to help maintain a robust understanding of online customer preferences and to enlighten the general purview of online customers' (e-shoppers') value proposition in online transactions. This will intend aid online providers to increase in online sales and customer retention as the core mandate of every company, and for that matter online providers in and beyond the Czech Republic.

Association rule mining was chosen for this study because of its tremendous and unique influence on research, helping studies retrieve the hidden and untapped knowledge embedded in a pool of data sets, particularly due to its segmenting and targeting approach in a large data set. The data mining algorithm used in this study demonstrates the way and manner that— despite the complex nature of online transactions—online providers can keep track of their regular customers and also lure more customers into their folds. The approach used, however, provides a clear picture and a gateway for online providers to succeed in their business endeavors and subsequently maximize their profit margin. However, the authors are lamenting on some limitations of the present study. The entire rule set that was generated and extracted was centered on some small aspect of online shopper-vendor interactions, notably price and specialized online shops. These form only a small part of the base of online interactions, and hence does not represent entire interactions that go on between the e-vendor and the e-shopper. Therefore, we cannot generalize our findings as a true reflection of what pertains to the online shopping environment. Different algorithms could be used or applied in the future to analyze and/or enhance the understanding of online shopper-vendor interactions, especially by blending them with other algorithms in different Countries and presenting multiple questions to the respondents.

Acknowledgement

References

[1]   D. Grewal, GR. Iyer, M. Levy: Internet retailing: enablers, limiters and market consequences, Journal of Business Research, 2004, 57 (7), pp. 703-713

[2]   C. Steinfield, A. Mahler, J. Bauer: Electronic commerce and the local merchant, Electronic Markets, 1999, 9 (1-2), pp. 51-57

[3] A. Hamidi, M. Safabakhsh: The impact of information technology on E. marketing, Procedia Computer Science, 2011, 3, pp. 365-368

[4] S. Min, M. Wolfinbarger: Market share, profit margin, and marketing efficiency of early movers, bricks and clicks, and specialists in e-commerce, Journal of Business Research, 2005, 58(8), pp. 1030-1039

[5] D. Brown: Searching for online gold, American Journalism Review, 2003, 25(5), pp. 54-61

[6] P. Ahluwalia, J. Hughes, V. Midha: Drivers of e-retailer peak sales period price behavior: an empirical analysis, International Journal of Accounting & Information Management, 2013, 21(1), pp. 72-90

[7] SJ. Lennon, KK. Johnson, KK, J. Lee: A perfect storm for consumer misbehavior: shopping on Black Friday, Clothing and Textiles Research Journal, 2011, 29 (2), pp. 119-134

[8] C. Topaloğlu: Consumer motivation and concern factors for online shopping in Turkey, Asian Academy of Management Journal, 2012, 17(2), pp. 1-19

[9] CH. Park, YG. Kim: Identifying key factors affecting consumer purchase behavior in an online shopping context, International Journal of Retail & Distribution Management, 2003, 31(1), pp. 16-29

[10] DJ. Reibstein: What attracts customers to online stores, and what keeps them coming back? Journal of the academy of Marketing Science, 2002, 30(4), pp 465-473

[11] M. Khalifa, V. Liu: Online consumer retention: Contingent effects of online shopping habit and online shopping experience, European Journal of Information Systems, 2007, 16(6), pp. 780-792

[12] J. C. Jiang, C. A. Chen, CA, C. C. Wang: Knowledge and trust in E-consumers' online shopping behavior. International Symposium on Electronic Commerce and Security, Guangzhou City, 2008, IEEE, pp. 652-656

[13] J. Kim, S. Forsythe: Factors affecting adoption of product virtualization technology for online consumer electronics shopping, International Journal of Retail & Distribution Management, 2010, 38(3), pp. 190-204

[14] S. Elliot, S. Fowell: Expectations versus reality: a snapshot of consumer experiences with Internet retailing, International Journal of Information Management, 2000, 20(5), pp. 323-336

[15] T. Bao, TLS. Chang: Why Amazon uses both the New York Times Best Seller List and customer reviews: an empirical study of multiplier effects on product sales from multiple earned media, Decision Support Systems, 2014, 67, pp. 1-8

[16]   K. Pousttchi, Y. Hufenbach: Engineering the value network of the customer interface and marketing in the data-rich retail environment, International Journal of Electronic Commerce, 2014, 18(4), pp. 17-42

[17]   Eurost: Ecommerce statistics explained, available at http://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce_statistics_for_individuals, download time: 22. 12. 2016

[18]   A.T Kearney: Global Retail E-Commerce Keeps On Clicking, available from www.atkearney.com, download time 06. 01. 2017

[19]   eMarketer: Worldwide retail ecommerce, available at www.emarketer.com/Article/ / Worldwide Retail Ecommerce Sales Will Reach $1.915 Trillion -by-2020/1011694, download time: 22. 12. 2016

[20]   T. Yang, R.T. Peterson: Customer Perceived Value, Satisfaction, and Loyalty: The Role of Switching Costs, Psychology & Marketing, 2004, 21, 10, pp. 799-822, ISSN: 07426046

[21]   Redakce: Mall, Alza.cz a ti druzí. Proč se z Česka stal ráj e-shopů. Deník Insider. available at http://www.denikinsider.cz/mall-alza-cz-a-ti-druzi-proc-se-z-ceska-stal-raj-e-shopu/?eid=630, download time: 28. 08. 2013

[22]   Czech Statistical Office: Individuals buying on the Internet in the EU countries, Iceland, Croatia and Norway, available at http://www.czso.cz/csu/redakce.nsf/i/nakupy_pres_Internet_v_zemich_eu/$File/nakupy_pres_Internet.xls, download time: 20. 07. 2010

[23]   Czech News Agency. Třetina elektroniky se v ČR prodá přes Internet, je to nejvíc v Evropě, Denik, available at http://www.denik.cz/ekonomika/tretina-elektroniky-se-v-cr-proda-pres-Internet-je-to-nejvic-v-evrope-20130810.html, download time: 15. 09. 2013

[24]   D. Hutchings, N. Best, M. Mahmuti: Global Perspective on Retail: Online Retailing, available at //annualreview.cushwake.com/downloads/01_Global_Perspective_on_Retail.pdf, download time: 22. 12. 2016

[25]   Ecommerce news: Ecommerce in the Czech Republic, available at https://ecommercenews.eu/ecommerce-in-the-czech-republic-grew-over-20-last-year/, download time: 22. 12. 2016

[26]   S. Brin, R. Motwani, C. Silverstein: Beyond market baskets: generalizing association rules to correlations, ACM SIGMOD Record, 1997, Vol. 26, No. 2, pp. 265-276

[27]   J. M. Adamo: Data mining for association rules and sequential patterns: sequential and parallel algorithms, Springer Science & Business Media, 2012

[28]   F. Schwenkreis: International Business Machines Corporation, Data mining for association rules and sequential patterns within data of inhomogeneous type, 2003, U.S. Patent 6,553,359

[29]   R. Agrawal, T. Imielinski, A. Swami: Database mining: a performance perspective, Transactions on Knowledge and Data Engineering, 1993, IEEE, 5 (6), pp. 914-925

[30]   R. Agrawal, R. Srikant: Fast algorithms for mining association rules, 20th Very Large Data Bases, 1994, Vol. 1215, pp. 487-499

[31]   C. Hidber: Online association rule mining, 1999, Vol. 28, No. 2, pp. 145-156

[32]   J. Vaidya, C. Clifton: Privacy preserving association rule mining in vertically partitioned data. 8th International Conference on Knowledge Discovery and Data Mining, 2002, pp. 639-644

[33]   C. Becquet, S. Blachon, B. Jeudy, J.F. Boulicaut, O. Gandrillon: Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data, Genome Biology, 2002, 3(12), research0067−1

[34]   C. Borgelt, M. R. Berthold: Mining molecular fragments: finding relevant substructures of molecules, International Conference on Data Mining, 2002, IEEE, pp. 51-58

[35]   M. J. Zaki, S. Parthasarathy, M. Ogihara, W. Li: New Algorithms for Fast Discovery of Association Rules, 3rd International Conference on Knowledge, Discovery and data Mining, 1997, Vol. 97, pp. 283-286

[36]   J. Han, J. Pei, M. Kamber: Data mining: concepts and techniques. Elsevier, 2011

[37]   M. Song, I.Y. Song, X. Hu, R.B. Allen: Integration of association rules and ontologies for semantic query expansion, Data & Knowledge Engineering, 2007, 63(1), pp. 63-75

[38]   Y. Chen, Q. Wang, J. Xie: Online social interactions: a natural experiment on word of mouth versus observational learning, Journal of Marketing Research, 2011, 48(2), pp. 238-254

[39]   A.G. Parsons: Non-functional motives for online shoppers: why we click, Journal of Consumer Marketing, 2002, 19(5), pp. 380-392

[40]   C. Zhang, S. Zhang: Association rule mining: models and algorithms, Springer-Verlag, 2002

[41]   X. Zhang, V. R. Prybutok, D. Strutton: Modeling influences on impulse purchasing behaviors during online marketing transactions, Journal of Marketing Theory and Practice, 2007, 15(1), pp. 79-89

[42]   G. Punj: Effect of consumer beliefs on online purchase behavior: the influence of demographic characteristics and consumption values, Journal of Interactive Marketing, 2011, 25(3), pp. 134-144

[43]   M. North: Data mining for the masses, Global Text Project, 2012, pp. 91-100